(21) Application No 9621066.1

(22) Date of Filing 09.10.1996

(71) Applicant(s)
Sony United Kingdom Limited

(Incorporated in the United Kingdom)

The Heights, Brooklands, WEYBRIDGE, Surrey,
KT13 0XW, United Kingdom

(72) Inventor(s)
Ahmad Sadjadian
Jonathan James Stone

(74) Agent and/or Address for Service
D Young & Co
21 New Fetter Lane, LONDON, EC4A 1DA,
United Kingdom

(51) INT CL^6
H04N 7/50

(52) UK CL (Edition P)
H4F FD3P FD3T FD30H FD30K FD30R FRG

(56) Documents Cited
EP 0727907 A1   EP 0637175 A2   EP 0588476 A2
US 5440344 A

(58) Field of Search
UK CL (Edition O) H4F FGM FGXX FRT FRW
INT CL^6 G06T 9/00 , H04N 5/926 7/30 7/34 7/36 7/50
Online: WPI,INSPEC

(54) Motion predicted image compression

(57)   A motion predictive inter frame compression system comprises a wavelet transform unit (80) for transforming time domain image data to the frequency domain. A motion estimator 30 operates in the time domain to produce motion vectors. The motion vectors are converted in a converter 31 to the frequency domain.
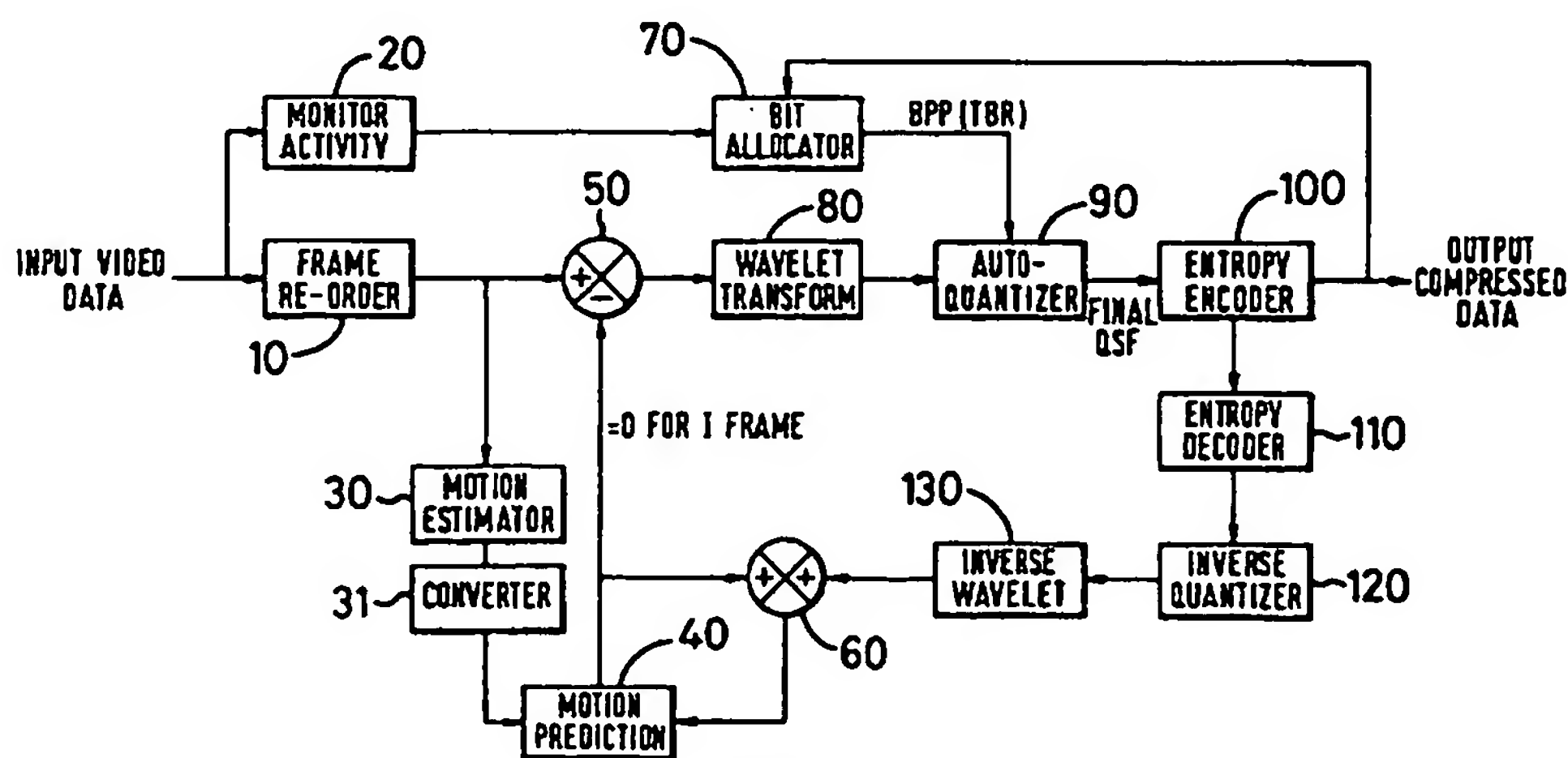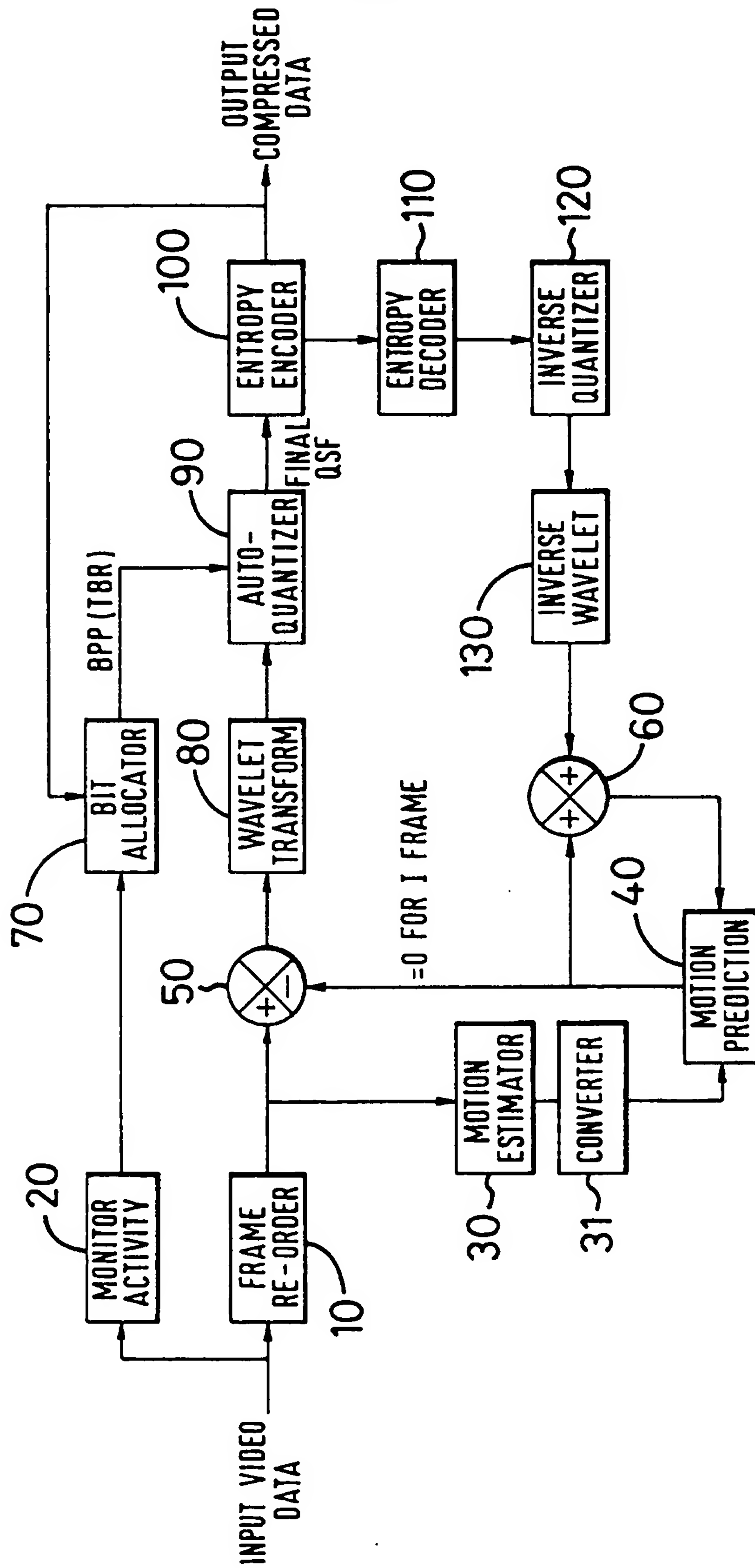
FIG.1
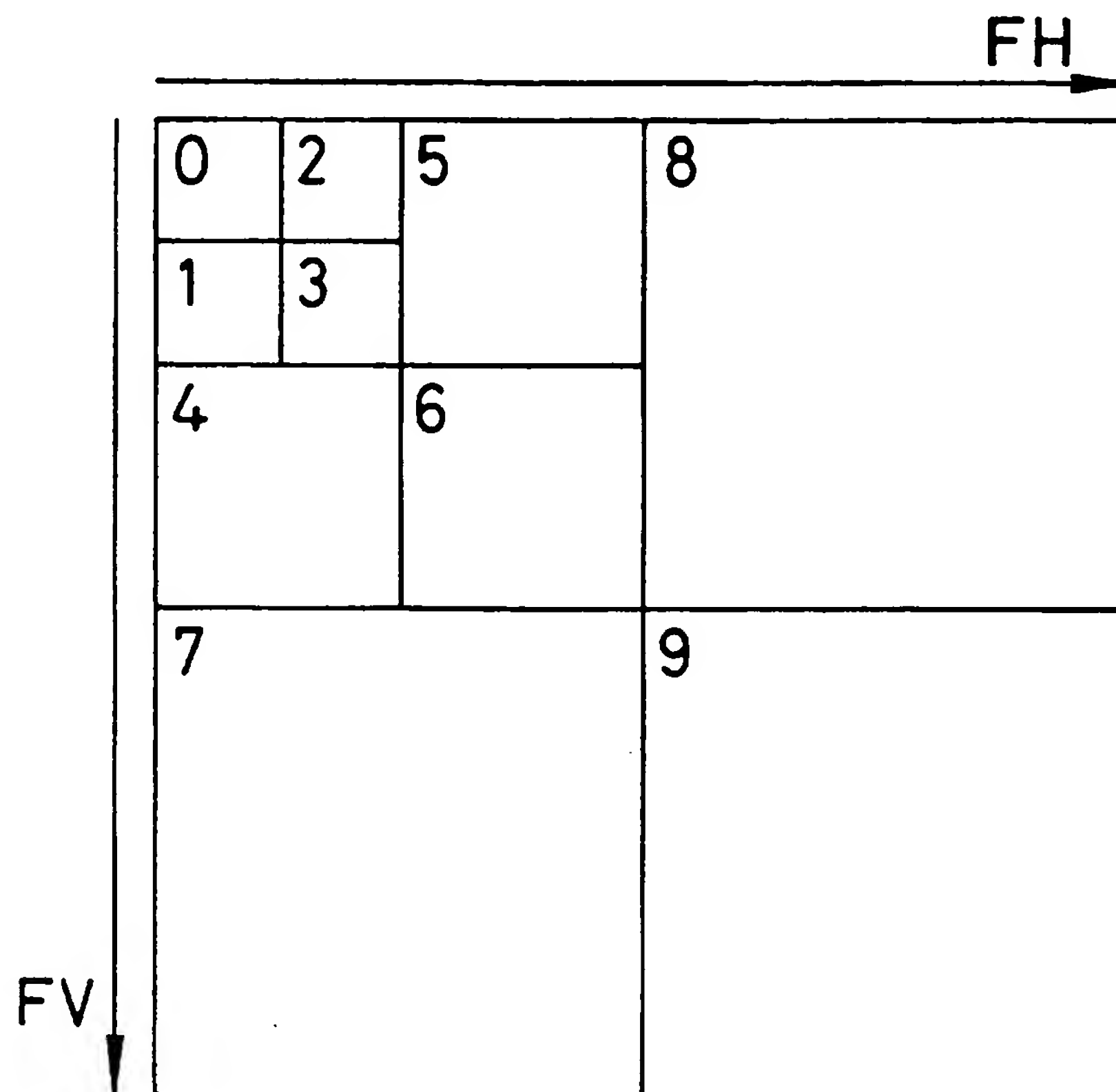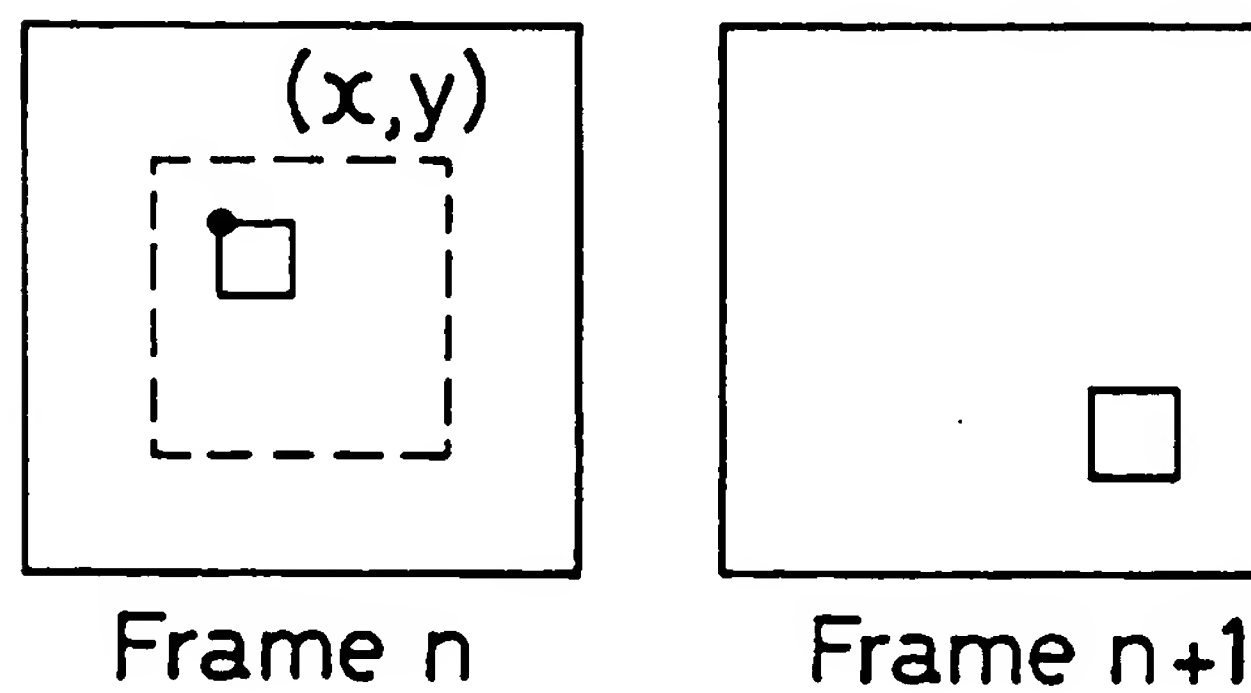
GB 2 318 244 A

FIG.1

FH

FV

0	2	5	8
1	3		
4		6	
7		9	

THREE STAGE WAVELET FREQUENCY DOMAIN REPRESENTATION

FIG.2

(x,y)

Frame n

Frame n+1

FIG.3

MOTION PREDICTED IMAGE SIGNAL COMPRESSION

The present invention relates to motion predicted image signal compression.

A motion predicted, inter frame, image signal compression system is known.

5 An example of such a system is MPEG-2 as set out in the Motion Pictures Expert Group II standard ISO/IEC Publication DIS 13818/2 "Information Technology - generic coding of motion pictures and associated audio information, March 1995. Such a system uses Discrete Cosine Transformation (DCT) of image data as one of several compression techniques.

10 To additionally compress the image data, motion prediction is used, in which there is calculated in the time domain the position in a reference frame of an image block, which image block (called the search block) occurs in a succeeding frame. That is done by comparing the search block with similar size blocks in the reference frame until a match (if any) is found. In place of the image information of the search

15 block, only the position of the matching block in the reference frame is used. The image information of the search block is then derived from the matching block in the reference frame. The matching block in the reference frame is used as a prediction of the search block. The position information so produced is termed a "motion vector".

20 It has been proposed to implement a motion-compensated inter-frame compression system using another known transform such a Wavelet Transform or a Sub-Band Transform in place of DCT.

When a Wavelet Transform or a Sub-Band Transform is used, an input image is transformed into two dimensional spatial frequency bands, each of which is a

25 differently sub-sampled version of the input image.

It has been proposed that frequency domain motion prediction is carried out on the frequency transformed data. To do that a search block is defined for each frequency band. For each frequency band the search block is compared to correspondingly sized blocks in a reference frame. Thus for each frequency

30 transformed image frame as many motion predictions are needed as there are frequency bands.

According to the present invention, there is provided a motion predicted inter-frame image signal compression system comprising means for transforming image data from one of a time domain and a frequency domain to the other, means operating in the said one domain to produce motion vectors and means for converting the said

5   motion vectors to the other domain.

Thus in an embodiment of the invention, image data is transformed from the time domain (the said one domain) to the frequency domain (the said other domain) by a Wavelet Transform. The motion prediction signals are produced in the time domain and are converted to the frequency domain.

10   This avoids using a plurality of search blocks for respective frequency bands.

A search block is used in the time domain to produce a motion vector. The motion vector is then converted to a set of motion vectors, one for each frequency band in the wavelet transformed image. In a preferred embodiment, the converting means converts the time domain vectors to vectors for each of the frequency bands

15   by scaling the time domain vectors proportionally to the sub-sampling factors of each of the blocks.

In the embodiment the frequency domain the motion vectors are produced from sub-sampled images which have lower resolution than a corresponding time domain image which is no sub-sampled. By producing motion vectors in the time

20   domain, the vectors are more accurate.

As a transform from one domain to another is a reversible process, motion vectors may be produced in e.g. the frequency domain and then converted to the time domain. It is believed that such conversion maybe advantageous where an image has been sub-sampled in the time domain before motion vector generation. The

25   conversion maybe achieved by suitably scaling-up a motion vector calculated from one of the wavelet sub-bands.

For a better understanding of the present invention, reference will now be made by way of example to the accompanying drawings, in which:

Figure 1 is a schematic diagram of a video data compression system; and

30   Figure 2 is a frequency domain representation of a video frame when transformed by a 3-stage wavelet transform.

Figure 1 is a schematic diagram of a video data compression apparatus comprising a frame reorderer 10, an activity detector 20, a motion estimator 30, a motion predictor 40, a subtracter 50, an adder 60, a bit allocator 70, a wavelet transform unit 80, an auto-quantiser 90, an entropy encoder 100, an entropy decoder 110, an inverse quantiser 120 and an inverse wavelet coder 130.

Many features of the apparatus of Figure 1 operate in a very similar manner to corresponding features of an MPEG encoder. Such features will not be described in detail here.

Typically, in an MPEG encoder, the video signal is divided into successive groups of pictures (GOPs). Within each GOP at least one picture is encoded as an "I-picture", or intra-picture, using only information present in that picture itself. This means that I-pictures can later be decoded without requiring information from other pictures, and so provide random entry points into the video sequence. However, the converse of this is that the encoding of I-pictures cannot make use of the similarity between successive pictures, and so the degree of data compression obtained with I-pictures is only moderate.

Further pictures within each GOP may be encoded as "P-pictures" or predicted pictures. P-pictures are encoded with respect to the nearest previous I-picture or P-picture, so that only the differences between a P-picture and the previous P- or I-picture needs to be transmitted. Also, motion compensation is used to encode the differences, so a much higher degree of compression is obtained than with I-pictures.

Finally, some of the pictures within a GOP may be encoded as "B-pictures" or bidirectional pictures. These are encoded with respect to two other pictures, namely the nearest previous I- or P-picture and the nearest following I- or P-picture. B-pictures are not used as references for encoding other pictures, so a still higher degree of compression can be used for B-pictures because any coding errors caused by the high compression will not be propagated to other pictures.

Therefore, in each GOP there are (up to) three classes of picture, I-, P- and B- pictures, which *tend to* achieve different degrees of compression and so *tend to* require different shares of the overall available encoded bit stream. Generally, I-pictures require a large share of the available transmission or storage capacity,

followed by P-pictures, and followed by B-pictures.

Briefly, therefore, the frame reorderer 10 receives input video data and acts on successive groups of pictures (GOP) to reorder the pictures so that each picture within the GOP is compressed *after* those pictures on which it depends. For example, if a B-picture (bi-directionally predicted picture) depends on a following I- or P-picture, it is reordered to be compressed after that I- or P- picture.

For example, if a GOP comprises the following four initial frames (in the order in which they are *displayed*), $I_0B_1B_2P_3...$, where the P-picture uses the I-picture as a reference and the two B- pictures use the surrounding I- and P-pictures as references, then the frame reorderer 10 will reorder the GOP to be *compressed* in the following order: $I_0P_3B_1B_2...$

I- pictures are intra-picture encoded, that is to say the encoding is not based on any other reference pictures. An I- picture in a GOP is therefore passed from the frame reorderer 10 to the wavelet transform unit 80, the auto quantiser 90 and the entropy encoder 100 to generate output compressed data representing that I- picture.

The compressed I-picture data is also passed from the entropy encoder 100 through a decompression chain formed by the entropy decoder 110, the inverse quantiser 120, and the inverse wavelet transform unit 130. This reconstructs a version of the I- picture present in the decoder which is passed to the motion predictor 40.

The next picture of the GOP to be compressed, which will generally be a P-picture which depends on the I- picture as a reference, is passed from the frame reorderer 10 to the motion estimator 30 which generates motion vectors indicative of image motion between the I- and P- pictures. The motion predictor 40 then generates a predicted version of the P picture using the motion vectors and the decoded version of the I- picture. This predicted version of the P- picture is subtracted from the actual P- picture by the subtracter 50 and the difference between the 2 frames is passed to the wavelet transform unit 80 for compression. As before, the encoded (compressed) difference data is output by the entropy encoder and is then decoded by the decompression chain 110,120,130 to regenerate a version of the difference data.

In the adder 60 the difference data is then added to the previously

decompressed version of the I- picture to generate a decompressed version of the P- picture which is then stored in the motion predictor 40 for use in the compression of the next picture.

This process continues, so that each picture which uses other pictures as a

5   reference is in fact compressed by encoding difference data between the input picture and a version of the input picture formed by motion prediction from a previously compressed and then decompressed version of the reference picture. This means that the compression is performed with respect to the pictures which will be available at the decompressor.

10   The activity detector 20 detects the image "activity", or "degree of detail" in blocks of each input image. This process will be described in more detail with reference to Figure 2 below.

The bit allocator 70 allocates target bit rates to whole pictures or blocks of the pictures in dependence on the image activity of pictures of the current GOP and the

15   degree of quantisation obtained for I-, B and P- pictures of the preceding GOP. In fact, the allocation can be made by allocating an overall target bit rate for each GOP (TBR_{GOP}) in proportions dependent on the actual quantity of data generated for the corresponding frame in the preceding GOP, or in accordance with the actual I:B:P ratio achieved with the preceding GOP. In this way, the allocation or the I:B:P ratio

20   can be "steered" to reflect the type of image content in use.

The target bits rates are supplied to the auto quantiser 90 which generates a suitable quantisation factor to be applied to the wavelet encoded data to comply with the target bit rates.

Referring to Figure 2, the wavelet transform unit 80 transforms time domain

25   video data to frequency domain data occupying, in this example, ten, two-dimensional spatial frequency sub-bands labelled 0 to 9 in Figure 2. In Figure 2 arrow FH indicates increasing horizontal frequency and arrow FV indicates increasing vertical frequency. The wavelet transform is known and thus will not be described in detail herein. Briefly, in the first stage of transformation the video data is sub-sampled

30   vertically and horizontally by a factor of 2 producing 4 sub-bands occupying in quadrants 7, 8 and 9 plus a sub-band in the upper left quadrant (0-6). The sub-bands

correspond to 1/4 size images.

The sub-band in the upper left quadrant containing the lowest frequencies of the four sub-bands is again transformed, again being sub-sampled horizontally and vertically by a factor 2 to produce 4 sub-bands 4, 5, 6 and the upper left quadrant (0,

5 1, 2, 3) each representing a 1/16 size image.

Again the sub-band in upper left quadrant is transformed to four sub-bands 0, 1, 2, 3, each representing a 1/64 size image.

The most significant image data tends to be present in sub-band 0 and the least significant in sub-band 9. The quantizer 90 quantises sub-band 0 with the greatest

10 accuracy and the sub-band 9 with the least accuracy to achieve compression without significant less of image quality.

Referring to Figure 3, the motion estimator 30 operates in the time domain to produce motion vectors. The motion estimator 30 is known and will not be described herein in detail. Briefly, an object in frame n positioned within a 16x16 pixel block

15 at position (x, y) defined by the top left hand corner of the block moves to a different position in the following frame n+1. Instead of encoding all the information of the block (the search block) in the frame n+1, only the position (x, y) of the same information in the preceding block in frame n is found by comparing the contents of the search block with the contents of one area (shown by the dotted line) around the

20 likely position of the block in frame n. It is assumed that in 1/25 or 1/30th of a second (depending on the frame rate), an object moving at a predetermined maximum speed will be within the search area.

Whilst the above description refers to only one search block, in practice all 16x16 blocks in the frame n+1 are compared with corresponding search areas in

25 frame n.

In accordance with the present invention, in the example shown in Figure 1, the time domain motion vectors produced by the estimator 30 are converted to the frequency domain in a converter 31.

In the example of Figure 1, the converter scales the motion vectors as follows,

30 it being assumed that the sub-bands 0 to 9 of the wavelet encoded video correspond to sub-sampled time domain images.

Referring to Figure 2, sub-bands 7, 8 and 9 are sub-sampled horizontally by 2 and vertically by 2.

Thus for motion vectors applicable to sub-bands 7, 8 and 9, the time domain vectors are divided by 2 horizontally and by 2 vertically.

5 Likewise for sub-bands 4, 5 and 6 which are sub-sampled by 4 horizontally and by 4 vertically, the time domain vectors are divided by 4 horizontally and vertically.

For sub-bands 0, 1, 2 and 3, the time domain vectors are divided by 8 horizontally and vertically.

10 The use of scaled time domain motion vectors with wavelet or sub-band transformed image data is especially advantageous. The filtering used for wavelet or sub-band filtering can be chosen to minimise aliasing between frequency bands. The converted time domain motion vectors provide greater accuracy than corresponding motion vectors derived in the frequency domain because the time domain motion 15 vectors are derived from a frame which twice, four times or eight times the resolution horizontally and vertically than the corresponding frequency domain frame, depending on which sub-band is being considered.

It will be appreciated that the converter 31 may be controlled to scale the time domain motion vectors according to the wavelet sub-band being processed. In MPEG 20 2, encoded signals, control information is carried in the bit stream indicating the type of data in the stream. The converter 31 would be controlled by the control information.

Alternatively, such control information maybe omitted or not used the converter operating according to a preset algorithm.

CLAIMS

1.   A motion predictive inter-frame image signal compression system, comprising means for transforming image data from one of a time domain and a frequency

5   domain to the other,

means operating in the said one domain to produce motion vectors, and

means for converting the motion vectors from said one domain to the other.

2.   A system according to claim 1, wherein the transforming means transforms

10   the image data from the time domain to the frequency domain.

3.   A system according to claim 2, wherein the image data is transformed into a plurality of spatial frequency bands, sub-sampled by respective factors.

15   4.   A system according to claim 3, wherein the converting means converts the time domain vectors to vectors for each of the frequency bands by scaling the time domain vectors proportionally to the sub-sampling factors of each of the blocks.

5.   A system according to claim 2, 3 or 4 wherein the transforming means

20   transforms the image data to the frequency domain by use of a wavelet transform.

6.   A motion predictive, inter-frame image signal compression system substantially as hereinbefore described with reference to the accompanying drawings.

| Application No: | GB 9621066.1 | Examiner: | Joe McCann |
| Claims searched: | All | Date of search: | 4 December 1996 |

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

    UK Cl (Ed.O): H4F(FGM,FGXX,FRT,FRW)

    Int Cl (Ed.6): H04N(5/926,7/30,7/34,7/36,7/50);G06T(9/00)

Other: Online: WPI, INSPEC

Documents considered to be relevant:

Category	Identity of document and relevant passage		Relevant to claims
X,Y	EP 0727907A1	(SAMSUNG) - see figure 17	X:1,2 Y:3,5
X,Y	EP 0637175A2	(MATSUSHITA) - see figure 3 and col 13, lines 10-12	X:1,2 Y:3,5
Y	EP 0588476A2	(SONY) - see abstract and figure 11	Y:3,5
X,Y	US 5440344	(YOSHINORI ET AL) - see fig 4 and col 10, lines 49-53	X:1,2 Y:3,5